What influences attendance at major American sporting events?

Parth Bansal, Nenad Berac, Taylor Furry, Thomas Sullivan, Ian Vogt

Presentation Outline

- Business Problem
- Data Description
  - Data Sources and Outline
  - Data Joining
- What we did
- What we found
- Key Takeaways/ Conclusion
- Appendix

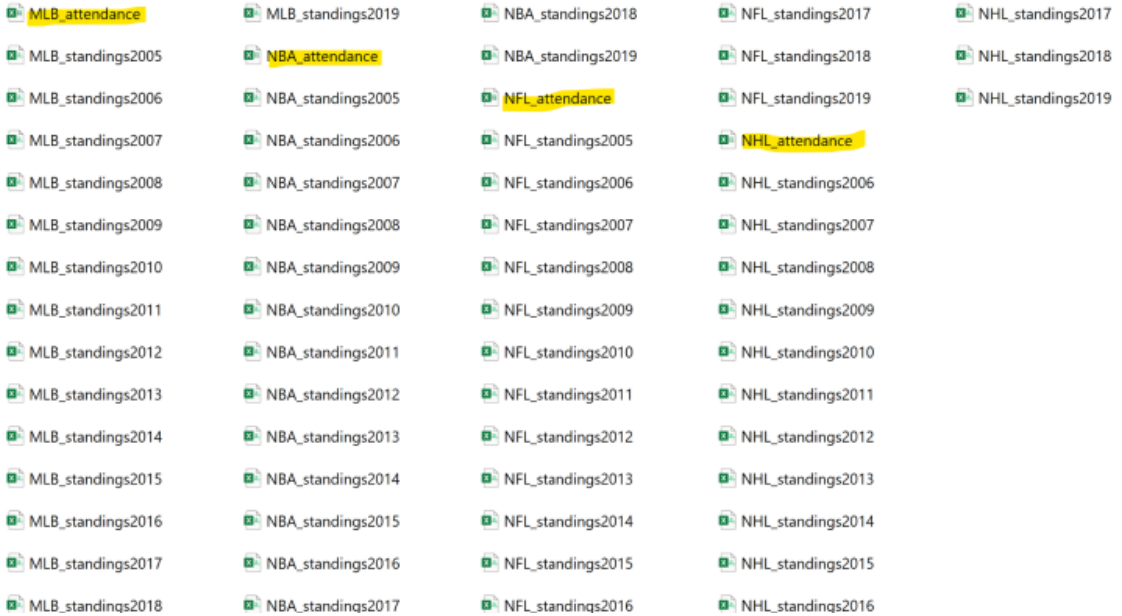
Business Problem

“Increase attendance at American sporting events by identifying and addressing the key factors that influence fans to attend games in-person.”

The objective is to determine which factors have the greatest impact on attendance and prioritize them in order to develop an effective strategy for increasing attendance and driving ticket sales, concessions, and on-site merchandising revenue for professional sports teams.

Data Sources and Outline

- Attendance data from ESPN (2010-2019), teams information from Sports Reference (2005-2019).
- Sports Reference data came year-by-year, ESPN (highlighted) all in one.
- Window of analysis is limited to pre-lockdown (2020).
- After the data is joined, each row represents 1 team season.



Data Joining

- Attendance data followed the same format across all 4 leagues, which made combining these into one dataframe an easy vertical stack.
- Standings data was formatted differently across leagues and included different variables. We wanted to pull a consistent set of variables from each league (wins, losses, ties) to be able to quantify team performance. This involved renaming columns and selecting only certain columns of interest. Each data frame was tagged with a year and league variable. A market variable was manually encoded.

Final Dataset

```
teams
```

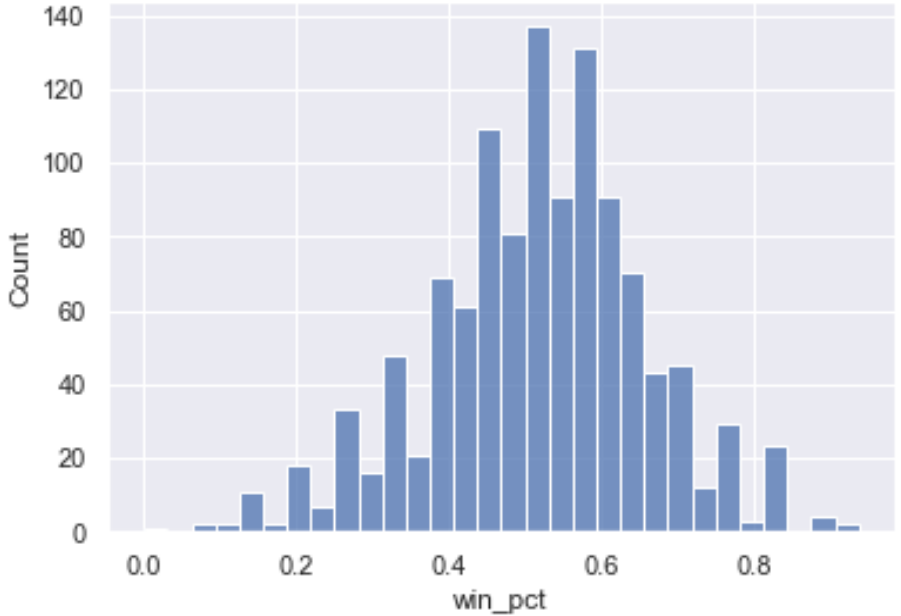
Out[18]:

	League	Year	Team	Home_Games	Home_Attendance	Home_Average	Home_Percentage	Team_Year	W	L	T_or_OTL	Market
0	MLB	2019	Los Angeles Dodgers	81	3974309	49065	71.0	2019 Los Angeles Dodgers	106	56	0	Los Angeles
1	MLB	2019	St. Louis Cardinals	81	3480393	42967	79.2	2019 St. Louis Cardinals	91	71	0	St. Louis
2	MLB	2019	New York Yankees	79	3304404	41827	66.6	2019 New York Yankees	103	59	0	Greater NYC
3	MLB	2019	Chicago Cubs	81	3094865	38208	75.2	2019 Chicago Cubs	84	78	0	Chicago
4	MLB	2019	Los Angeles Angels	81	3023010	37321	66.6	2019 Los Angeles Angels	72	90	0	Anaheim
...	...	...	...	...	...	...	...	...	...	...	...	...
1157	NHL	2012	Anaheim Ducks	40	591371	14784	0.861	2012 Anaheim Ducks	34	36	12	Anaheim
1158	NHL	2012	Columbus Blue Jackets	41	601061	14660	0.801	2012 Columbus Blue Jackets	29	46	7	Columbus
1159	NHL	2012	Dallas Stars	41	583306	14227	0.768	2012 Dallas Stars	42	35	5	Dallas
1160	NHL	2012	New York Islanders	41	540838	13191	0.813	2012 New York Islanders	34	37	11	Greater NYC
1161	NHL	2012	Arizona Coyotes	41	509241	12421	0.723	2012 Arizona Coyotes	42	27	13	Phoenix

1162 rows × 12 columns

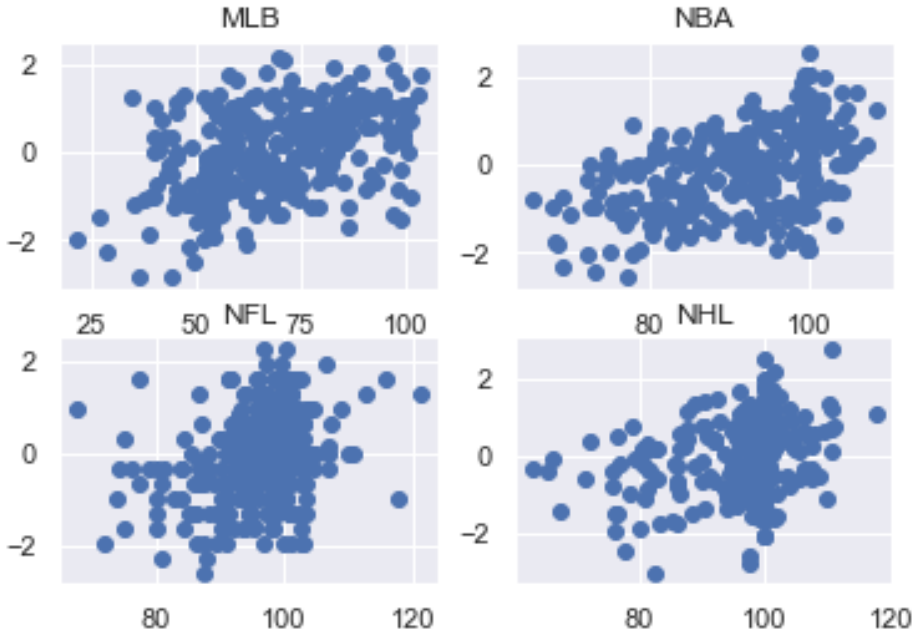
What We Did

- Created a win percentage variable and standardized by league



What We Did

- Standardized Home Percentage by league

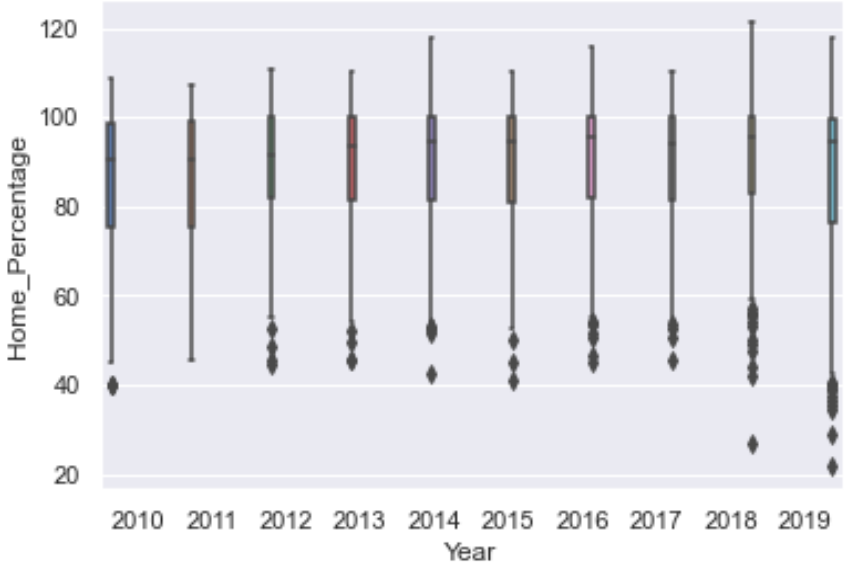


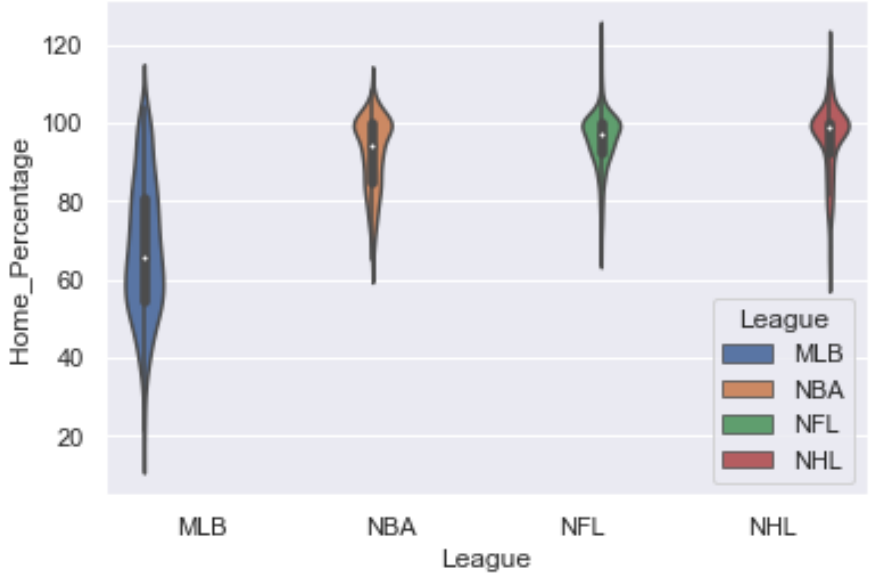

```
Overall correlation between Home_Percentage and win_pct is: 0.26058884769388174
MLB correlation between Home_Percentage and win_pct is: 0.41837080607190696
NBA correlation between Home_Percentage and win_pct is: 0.5206767101614531
NFL correlation between Home_Percentage and win_pct is: 0.3135772322703939
NHL correlation between Home_Percentage and win_pct is: 0.28751516008515865
```

What We Did

- Analyzed attendance by year, market, and league

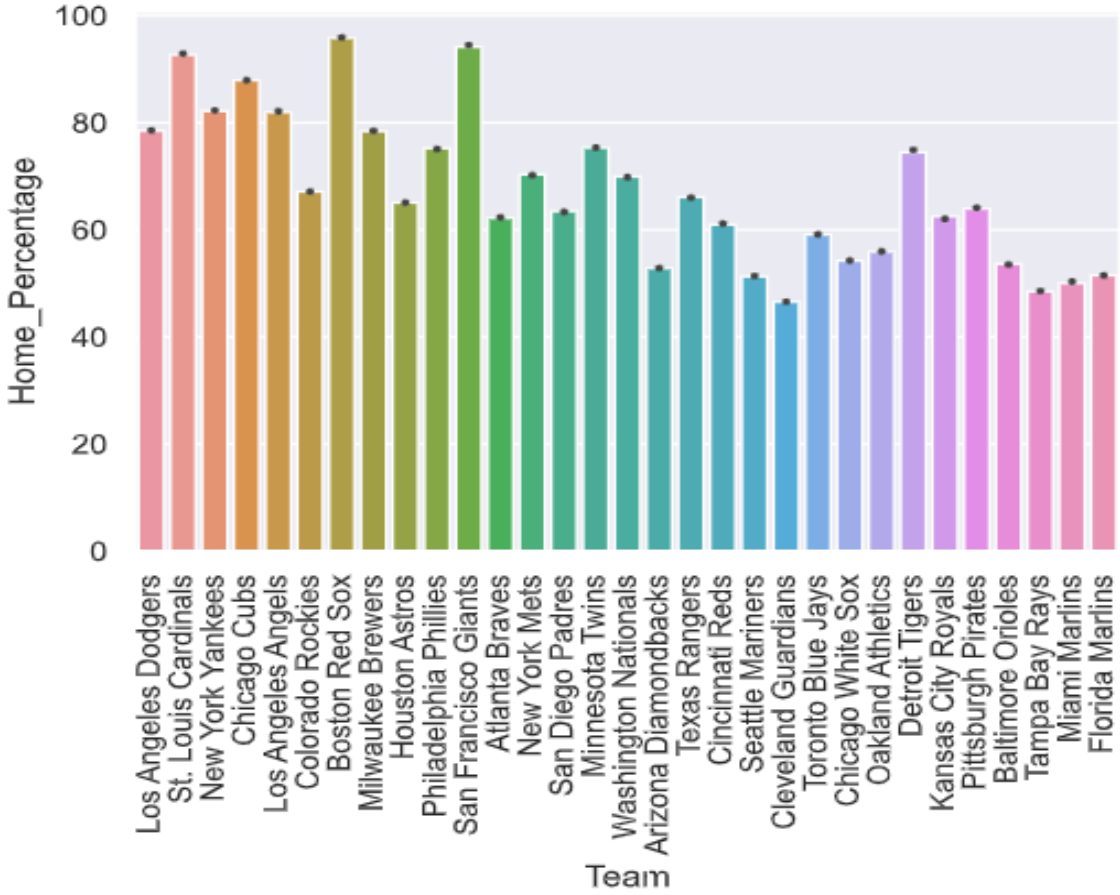
Market	count	mean	std	min	25%	50%	75%	max
Las Vegas	2.0	104.700000	1.131371	103.9	104.300	104.70	105.100	105.5
Winnipeg	8.0	101.462500	2.596667	100.0	100.000	100.10	101.950	107.5
Montreal	8.0	99.912500	0.412094	98.9	100.000	100.05	100.100	100.1
Oklahoma City	10.0	99.860000	0.350238	98.9	100.000	100.00	100.000	100.0





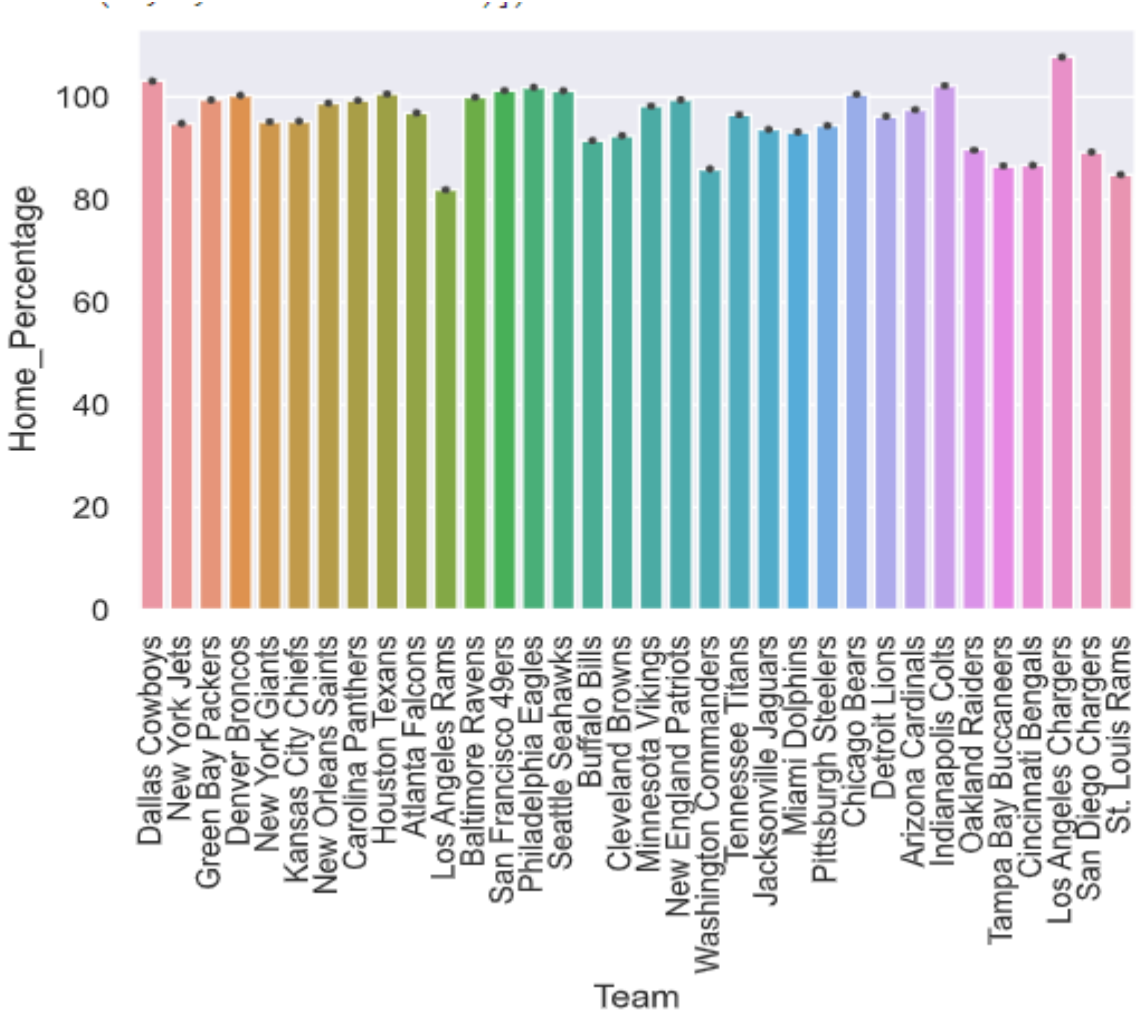
What We Did

- Attendance by team in the MLB



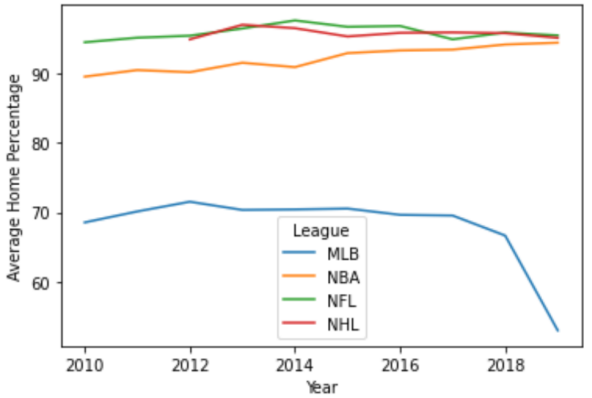
What We Did

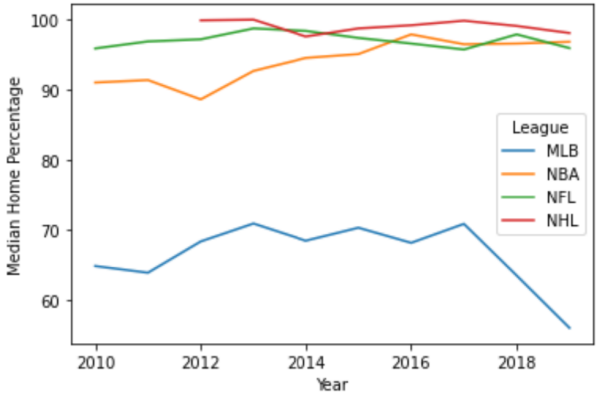
- Attendance by team in the NFL



What We Found

- Analysis observes mean and median home capacity percentage over time by league across the data.
- MLB games have far worse mean and median ticket sales across the entire window of analysis, with a significant drop-off in 2018 and 2019.
- The other three leagues of interest maintain mean and median rates around or above 90% across the entire sample.





What We Found

- Markets with the highest average attendance are all ones with counts less than or equal to 10.
- Two highest attended markets, Las Vegas and Winnipeg, are both teams that joined the NHL within the last 10 years
- Out of all the cities that have multiple sports teams, Boston and the Bay Area (San Francisco and San Jose) have the highest attendance rates.
- Boston stands out for sustaining high attendance rates across lots of iterations over the data (38 team seasons).

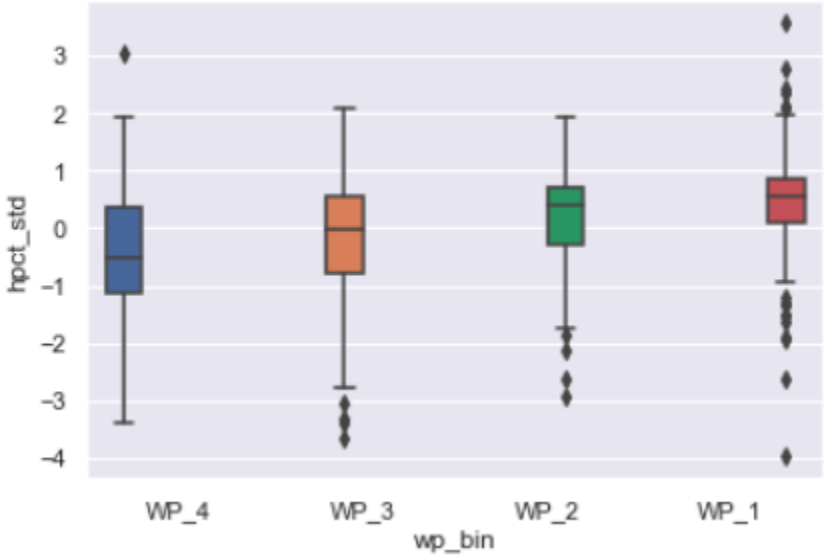
Market and Attendance

```
df.groupby(['Market'])['Home_Percentage'].describe().sort_values(by='mean', ascending=False)
```

Market	count	mean	std	min	25%	50%	75%	max
Las Vegas	2.0	104.700000	1.131371	103.9	104.300	104.70	105.100	105.5
Winnipeg	8.0	101.462500	2.596667	100.0	100.000	100.10	101.950	107.5
Montreal	8.0	99.912500	0.412094	98.9	100.000	100.05	100.100	100.1
Oklahoma City	10.0	99.860000	0.350238	98.9	100.000	100.00	100.000	100.0
Portland	10.0	99.800000	3.092284	94.0	99.450	100.05	102.450	102.7
Green Bay	10.0	99.450000	5.210939	95.6	95.925	96.70	104.375	107.2
Edmonton	8.0	99.387500	0.818426	98.4	98.400	99.95	100.000	100.0
Vancouver	8.0	99.087500	3.752118	95.3	97.025	98.40	100.125	107.2
San Antonio	10.0	98.940000	0.693141	97.4	98.700	99.05	99.200	100.1
San Jose	8.0	98.687500	1.580631	95.4	98.125	99.25	99.775	100.0
Calgary	8.0	98.675000	1.543419	95.9	97.775	99.15	100.000	100.1
Boston	38.0	98.402632	3.902285	78.0	97.300	100.00	100.000	101.7
Nashville	18.0	97.772222	3.337806	90.1	95.500	99.20	100.000	101.9
San Francisco	20.0	97.680000	8.404673	64.6	99.175	99.30	101.275	103.4
Salt Lake City	10.0	95.780000	3.264897	90.0	94.000	97.00	97.825	100.0
New Orleans	20.0	95.265000	5.310494	81.8	93.850	95.80	99.950	100.2
Orlando	10.0	94.880000	5.326412	86.2	93.125	93.80	98.725	102.6
Indianapolis	20.0	94.625000	9.818665	74.5	91.025	95.50	103.575	106.3

What We Found

- After we standardized win_pct, we grouped each team's winning percentage for that season into 4 bins.
- WP_1 indicated they were in the top quartile of winning percentage while WP_4 indicated that they were in the bottom quartile of winning percentage.
- Teams with the highest winning percentages in their leagues, on average, have better home attendance



Key Takeaways

- Team Winning% has a moderate, positive correlation with attendance rates, this correlation is strongest in the NBA and MLB
    - The Philadelphia 76ers were a team whose attendance rates were strongly correlated with Team Win % (r=0.80)
- The MLB had much lower average attendance rates compared to the other 3 major leagues
    - Average attendance rates continuously decreased in the final 3 years
    - While Certain MLB teams like the Red Sox and Giants maintained very high attendance rates, ⅓ of MLB teams averaged less than 60% Home attendance
- The NHL and NFL had relatively steady, high attendance rates, NBA attendance rates increased throughout the decade

Key Takeaways

- Markets with only one major professional team had the highest attendance rates overall
  - The top 9 best attended markets had only one major professional team
  - The San Antonio and Green Bay saw consistent high attendance rates throughout the decade
- Out of the top 4 highest attended markets, 3 of them were new markets to gain a professional team (Las Vegas, Winnipeg, OKC)
  - Las Vegas was the market with the highest average attendance overall
- For markets with multiple sports teams, Boston had the h and most consistent attendance rates

Potential opportunities to expand project

- Analyze Covid-19 impact on attendance rates
  - Have certain sports/markets rebounded from the pandemic better than others?
- Add variable to account for star power of a team
  - In the NBA, superstars like LeBron James can directly influence the attendance for a team
- Incorporate data to see how attendance changes throughout the course of the season
  - Teams out of playoff contention often see decline in attendance
- Deeper dive on MLB
  - Worrying trends from an ownership perspective
  - Significant decrease in attendance during the final 3 years of study

Appendix

Data dictionary:

- `League` : Categorical: The professional sports league that each team belongs to. Values are MLB, NFL, NBA, NHL
- `Year` : Int: Years go from 2010-2019
- `Team` : Categorical: The team. 30 possibilities for MLB and NBA, and 32 for NFL and NHL
- `Home_Games` : Int: The number of home games throughout the year. Not more than 82 for MLB, 50 for NBA and NHL and 10 for NFL. Low values indicate teams have just moved to new stadium
- `Home_Attendance` : Int: Total number of people that came to home games for that given year. Depending on the league, can be in the millions. Not more than 4,000,000 in our dataset
- `Home_Average` : Int: Average amount of people at each home game. Below 20,000 for NBA and NHL. Up to 50,000 or 60,000 for MLB, NFL
- `Home_Percentage` : Float: Home_Average divided by capacity of the stadium. Can be more than 100 if a team oversold their tickets. Generally not less than 20
- `Team_Year` : Categorical: Combines Year and Team column
- `W` : Int: The number of wins in that season. Not more than 108 for MLB, 16 for NFL, 70 for NBA and NHL
- `L` : Int: The number of losses in that season. Generally cannot be greater than 115 for MLB, 16 for NFL, 70 for NBA and NHL
- `T_or_OTL` : Float: The number of ties in that season. Between 0 and 18. Always 0 for NBA and MLB, generally not more than 3 for NFL
- `Market` : Categorical: A variable that accounts for the city or area the team is in. Accounts for the fact that some cities have more than one team and maybe even more than one team in a specific league